

## PATENT COOPERATION TREATY

PCT

## NOTIFICATION OF ELECTION

(PCT Rule 61.2)

From the INTERNATIONAL BUREAU

To:

United States Patent and Trademark  
Office  
(Box PCT)  
Crystal Plaza 2  
Washington, DC 20231  
ÉTATS-UNIS D'AMÉRIQUE

in its capacity as elected Office

Date of mailing (day/month/year)  
01 April 1999 (01.04.99)

International application No.  
PCT/GB98/02316

Applicant's or agent's file reference  
SP/1796 PCT

International filing date (day/month/year)  
31 July 1998 (31.07.98)

Priority date (day/month/year)  
31 July 1997 (31.07.97)

## Applicant

DAVID, William, Ian, Fraser et al

## 1. The designated Office is hereby notified of its election made:

☒ in the demand filed with the International Preliminary Examining Authority on:  
12 February 1999 (12.02.99)

☐ in a notice effecting later election filed with the International Bureau on:

2. The election ☒ was  
☐ was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO  
34, chemin des Colombettes  
1211 Geneva 20, Switzerland

Facsimile No.: (41-22) 740.14.35

Authorized officer

Lazar Joseph Panakal

Telephone No.: (41-22) 338.83.38

REC'D 28 MAY 1999

WIPO PCT

## INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference SP/1796 PCT	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416) <b>FOR FURTHER ACTION</b>	
International application No. PCT/GB98/02316	International filing date (day/month/year) 31/07/1998	Priority date (day/month/year) 31/07/1997
International Patent Classification (IPC) or national classification and IPC G01N23/20		
Applicant COUNCIL FOR THE CENTRAL LABORATORY OF THE ..et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.



2. This REPORT consists of a total of 4 sheets, including this cover sheet.

- ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

- I ☒ Basis of the report
- II ☐ Priority
- III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- IV ☐ Lack of unity of invention
- V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- VI ☒ Certain documents cited
- VII ☐ Certain defects in the international application
- VIII ☐ Certain observations on the international application

Date of submission of the demand 12/02/1999	Date of completion of this report 21.05.99
Name and mailing address of the international preliminary examining authority:  European Patent Office D-80298 Munich Tel. (+49-89) 2399-0 Tx: 523656 epmu d Fax: (+49-89) 2399-4465	Authorized officer Papantoniou, E Telephone No. (+49-89) 2399 

**INTERNATIONAL PRELIMINARY  
EXAMINATION REPORT**

International application No. PCT/GB98/02316

**I. Basis of the report**

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

**Description, pages:**

1-18 as originally filed

**Claims, No.:**

1-20 as originally filed

**Drawings, sheets:**

1/15-15/15 as originally filed

2. The amendments have resulted in the cancellation of:

- ☐ the description, pages:  
☐ the claims, Nos.:  
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

**INTERNATIONAL PRELIMINARY  
EXAMINATION REPORT**

International application No. PCT/GB98/02316

**V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

**1. Statement**

Novelty (N)	Yes:	Claims	1 - 20
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1 - 20
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1 - 20
	No:	Claims	

**2. Citations and explanations**

**see separate sheet**

**VI. Certain documents cited**

**1. Certain published documents (Rule 70.10)**

and / or

**2. Non-written disclosures (Rule 70.9)**

**see separate sheet**

**Re Item V**

**Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

The subject matter of independent claims 1 and 12 defines a method/apparatus used for determining the crystal structures of powder samples using a structure factor analyser for generating a reduced data representation of the sample.

As mentioned in the present application (pages 1 and 2), analysis of powder samples is hampered by the fact that the diffraction pattern may include overlapping reflections making it difficult to identify the crystalline structure. The known prior art overcomes this problem by using a time consuming point to point comparison with known crystal structure models.

Object of the present application is to overcome this problem. This objective is achieved with the use of the factor analyser which uses a generic algorithm for eliminating data which are not covariant, as defined in present claims 1 and 12.

Such solution is not known in the prior art discussed in the present application and thus claims 1 and 12 meet the requirements of Article 33 PCT.

**Re Item VI**

**Certain documents cited**

The claimed priority of the present application could not be verified. It is noted that if this priority is not valid, then the intermediate documents:

- K. SHANKLAND: "Crystal structure determination from powder diffraction data by the application of a genetic algorithm." ZEITSCHRIFT FÜR PHYSIK, vol. 212, no. 8, 8 August 1997, pages 550-552, and
- B. M. KARIUKI ET AL.: "The application of a genetic algorithm for solving crystal structures from powder diffraction data." CHEMICAL PHYSICS LETTERS, vol. 280, no. 3,4, 5 December 1997, pages 189-195.

could be used against the novelty and inventive step of the present claims.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>G01N 23/20</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/06824</b> <b>(43) International Publication Date:</b> 11 February 1999 (11.02.99)
<b>(21) International Application Number:</b> PCT/GB98/02316 <b>(22) International Filing Date:</b> 31 July 1998 (31.07.98)  <b>(30) Priority Data:</b> 9716278.8                      31 July 1997 (31.07.97)                      GB  <b>(71) Applicant (for all designated States except US):</b> COUNCIL FOR THE CENTRAL LABORATORY OF THE RESEARCH COUNCILS [GB/GB]; Chilton, Didcot, Oxfordshire OX11 0QX (GB).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DAVID, William, Ian, Fraser [GB/GB]; 24 Appleford Drive, Abingdon, Oxfordshire OX14 2BZ (GB). SHANKLAND, Kenneth [GB/GB]; 83a The Vineyard, Abingdon, Oxford OX14 3PG (GB).  <b>(74) Agent:</b> PERKINS, Sarah; Stevens, Hewlett & Perkins, 1 Serjeants' Inn, Fleet Street, London EC4Y 1LL (GB).		<b>(81) Designated States:</b> JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> METHOD AND APPARATUS FOR DETERMINING MOLECULAR CRYSTAL STRUCTURES		
<b>(57) Abstract</b>  An improved method and apparatus for determining molecular crystal structures is provided which enables molecular crystal structures to be identified using only powder diffraction data, that is considerably faster than conventional crystal modelling techniques. With the improved crystal modelling method and apparatus trial molecular crystal structures are reduced to a unique set of variables based on co-ordinates representative of the location and orientation of the molecule and where appropriate at least one co-ordinate representative of a torsion angle, bond length or bond angle. In addition, the total quantity of experimental powder diffraction data is reduced to provide a reduced representation of the diffraction data in the form of a structure factor intensity listing and covariance matrix. Trial crystal structures are postulated and each is defined using the set of variables which are used in determining a fitness of each trial structure with respect to the reduced representation of the experimental data. A trial crystal structure is output as an accurate representation of the actual crystal structure, when the fitness value for the trial structure is less than or equal to a predetermined threshold value. With the improved crystal modelling method and apparatus, identification of complex molecular crystal structures can be performed in seconds or minutes using the current generation of conventional personal computers or workstations as opposed to the hours and often days required with conventional techniques.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## METHOD AND APPARATUS FOR DETERMINING MOLECULAR CRYSTAL STRUCTURES

The present invention provides an improved method and apparatus  
5 for determining molecular crystal structures from powder diffraction data.  
In particular, the present invention enables molecular crystal structures to  
be identified using only powder diffraction data in a manner that is  
considerably faster than is currently the case. Furthermore, with the  
present invention the molecular crystal structure of large organic  
10 molecules, such as pharmaceutical compounds, can be determined using  
data from powder diffraction analysis.

Information on the molecular crystal structure of a molecule is  
usually obtained through irradiation of a single crystal of the molecule with  
neutrons or X-rays. Subsequent analysis of the resultant diffraction  
15 pattern, which consists of a series of angularly spaced intensity peaks with  
each peak representing an individual Bragg reflection, provides information  
on the structure. Whilst this single crystal diffraction technique is an  
effective technique for determining the crystal structure of a molecule, it  
can often prove difficult to grow the single crystals necessary for the  
20 analysis. Moreover, where the molecule can crystallise in more than one  
polymorphic form, it is sometimes the case that it can prove very difficult to  
grow a single crystal of a particular polymorph.

To address these problems, a powder diffraction analysis was  
developed in which a crystalline powder of the material under analysis is  
25 irradiated instead of a single crystal. Analysis of the resultant diffraction  
pattern is hampered by the fact that the diffraction pattern may include  
Bragg reflections that partially or fully overlap one another, making it  
difficult for individual reflections to be identified, and their associated  
intensities quantified. An example of experimental data from irradiation of  
30 a powder sample of the drug substance cimetidine in the form of a graph  
representing intensity of the Bragg reflections with respect to angular  
position is shown in Figure 1. Currently, in order to identify molecular



crystal structures diffraction patterns of this type are used in a point-to-point comparison with diffraction data calculated from a postulated model of the crystal structure. If there is good agreement between the measured and calculated diffraction data, it may be assumed that the postulated  
5 structure is close to the true crystal structure of the molecule. In general, good agreement is only obtained when there is significant prior knowledge of the true crystal structure of the molecule, as there is an infinite number of crystal structures that may be postulated and compared to the experimental data. Moreover, such analysis is very slow and time  
10 consuming because of the extremely large body of point-to-point data that must be compared for each postulated crystal model in turn.

The present invention seeks to address the problems discussed above with respect to existing diffraction analysis techniques and in particular to provide a method and apparatus that is considerably faster  
15 than conventional techniques. The present invention further seeks to provide a method and apparatus for determining molecular crystal structures which employs data obtained from the irradiation of crystalline powders and permits analysis without the need for prior knowledge of any approximate crystal structure.

20 The present invention provides a method for determining molecular crystal structures from powder diffraction data comprising the steps of: generating a reduced representation of the powder diffraction pattern in dependence on a predetermined unit cell and space group of the molecule under examination in which the total quantity of diffraction data is  
25 significantly reduced whilst maintaining the characteristics of the diffraction data that are representative of the crystal structure under examination; determining a set of variables for describing trial molecular structures, derived from predetermined internal co-ordinates and said space group; assigning values to said variables thereby creating a population of trial  
30 structures each defined by a unique set of values for said variables; calculating a fitness for each trial structure with respect to the reduced representation of the powder diffraction pattern; determining whether any

one of the calculated fitnesses is less than or equal to a predetermined threshold; where none of the calculated fitnesses is less than or equal to the threshold value, selecting at least one survivor from the population of trial structures, altering the values of the variables of at least one of the survivors in accordance with one or more predetermined rules, calculating the fitnesses of the new trial structures; and repeating the steps of selecting survivors, altering the values of the variables and calculating the fitnesses of the new trial structures until at least one of the calculated fitnesses is less than or equal to the threshold value, and where at least one of the calculated fitnesses is less than or equal to the threshold, outputting at least one trial molecular crystal structure represented by the successful sets of values.

In a further aspect the present invention provides apparatus for determining molecular crystal structures comprising a structure factor analyser for generating from experimental powder diffraction data for the molecule under examination a reduced representation of the powder diffraction pattern based on a predetermined unit cell and space group in which the total quantity of diffraction data is significantly reduced whilst the characteristics of the diffraction data representative of the crystal structure under examination are maintained; a controller for determining a set of variables for describing trial molecular structures, derived from predetermined internal co-ordinates and said space group; a searching processor for creating a population of trial structures each defined by a unique set of values for said variables said searching processor including a fitness analyser for calculating a fitness for each trial structure with respect to the reduced representation of the powder diffraction pattern, a thresholding device for determining whether any one of the calculated fitnesses is less than or equal to a predetermined threshold, a survivor selector for selecting at least one survivor from the population of trial structures, a variable adjustment device for altering the values of the variables of at least one of the survivors and output means for outputting the one or more trial molecular crystal structures having calculated

fitnesses less than or equal to the threshold value.

The reduced representation preferably consists of identification of each reflection in the powder diffraction data along with associated weighting factors and ideally is in the form of a structure factor intensity listing and associated covariance matrix. Reference herein to reflections is intended as reference to individual Bragg reflections or peaks in the diffraction data resulting from reflections of the incident radiation from the structure of the molecule.

Moreover, preferably, the fitness  $\chi^2$  of each of the trial structures is determined using the following function:

$$\chi^2 = \sum_h \sum_k \{ (I_h - c|F_h|^2) (V^{-1})_{hk} (I_k - c|F_k|^2) \}$$

where:

$I_{h,k}$  = extracted intensity from the structure factor analyser

$V_{hk}$  = covariance matrix from the structure factor analyser

$c$  = a scale factor

$F_{h,k}$  = calculated structure factor from trial structure

The plurality of co-ordinates preferably consist of three co-ordinates representative of the location of the molecule within the unit cell, three co-ordinates representative of the orientation of the molecule within the unit cell and one or more co-ordinates representative of respective torsion angles.

In a preferred embodiment the structure factor analyser additionally automatically determines the optimal unit cell and space group for the molecule under examination instead of the unit cell and space group being predetermined manually and input into the structure factor analyser. More preferably a co-ordinate generator is provided for automatically determining the set of internal co-ordinates instead of the set of internal co-ordinates being predetermined manually and input into the controller.

The search for a three dimensional structure of a molecule which

would produce a powder diffraction pattern nearly identical to available experimental results is greatly simplified with the present invention and considerably speeded up by a factor of typically up to 1000 or more. This is achieved by the reduction of trial molecular crystal structures to a unique  
5 set of variables that does not explicitly include the co-ordinates of individual atoms in the molecule but instead includes a single set of co-ordinates for the location and orientation of the entire molecule. Reduction of the experimental powder diffraction data to data representative of each reflection in combination with a weighting factor also assists in speeding up  
10 the analysis of trial crystal structures in which the fitness of each trial structure with respect to the reduced experimental data is determined using structure factor analysis.

The present invention relies on the fact that at its most basic, a molecular crystal structure can be represented by a set of internal co-  
15 ordinates describing the molecule under investigation together with co-ordinates describing the location and orientation of the molecule within a unit cell of which only some but not all need be variable. The reduction of the molecular crystal structure to such a set of variables enables analysis of the trial structures to be performed much more quickly than an analysis  
20 performed using the conventional method of describing the crystal structure in terms of the fractional or Cartesian co-ordinates of every atom in the asymmetric unit of the structure. Such conventional representations are considered to be unworkable in a model building sense because of the computing power necessary to position individual atoms independently of  
25 each other.

The representation of the trial structures used in the invention along with the novel fitness function means that analyses can be performed in seconds or minutes on the current generation of conventional personal computers or workstations. Moreover, the representation is versatile as it  
30 allows the invention to work with flexible molecules as well as multiple fragments.

An embodiment of the present invention will now be described by

way of example with reference to the accompanying drawings, in which:

Figure 1 is a graph of experimental data from x-ray powder diffraction analysis of cimetidine showing 315 reflections, using an irradiation wavelength of 1.5285Å and a data range for 2θ of 8°-56°;

5        Figure 2 is a schematic representation of the 2D molecular structure of cimetidine;

Figure 3 is a flow diagram of the method steps for determining a molecular structure in accordance with the present invention;

Figure 4 is a diagram of the crystal structure of cimetidine;

10       Figures 5a, 5b, 5c and 5d are diagrams showing the progressive development of a trial crystal structure for cimetidine, employing the method and apparatus in accordance with the present invention, overlying the diagram of Figure 4;

Figure 6 is a graph showing the fitness of a trial crystal structure for cimetidine with respect to generations, employing the method and apparatus in accordance with the present invention;

Figures 7a, 7b and 7c show the molecular structure of dopamine deuterobromide, a graph of the development of a trial structure for the crystal and a diagram of the solution respectively, employing the method and apparatus of the present invention;

Figure 8 is a graph of experimental data from x-ray powder diffraction analysis of capsaicin using an irradiation wavelength of 0.6528Å and a data range for 2θ of 2.7°-22.5°;

Figure 9 is a tabulation of part of the reduced representation of the powder diffraction data of Figure 8 generated in accordance with the method of the present invention; and

Figure 10 is a diagram comparing the crystal structure of capsaicin obtained from single crystal diffraction data with the crystal structure obtained using powder diffraction data alone in a method in accordance with the present invention.

The present invention will be described with reference to an experimental determination of the crystal structure of the molecule

cimetidine, a histamine H<sub>2</sub> antagonist used in the treatment of stomach ulcers, for which a full single crystal structure (monoclinic Form A) determination has already been performed. Figure 2 shows the 2D chemical formula of the cimetidine molecule, whilst the known arrangement of the cimetidine molecules within the unit cell of the crystal structure is shown in Figure 4.

To determine the molecular crystal structure of cimetidine employing the method and apparatus of the present invention with reference to Figure 3, initially a conventional powder diffraction pattern (10) is obtained from a crystalline powder sample of cimetidine. The resultant diffraction pattern is shown in Figure 1. The experimental diffraction data (10) is input into a cell dimension analyser (12). The cell dimension analyser (12) uses conventional techniques to assess the diffraction pattern in order to determine the unit cell dimensions of the crystal structure. The generation of the unit cell dimensions may alternatively be performed manually, however, it is preferred that the unit cell dimensions be generated automatically using the crystal modelling apparatus. The diffraction pattern is also input to a structure factor analyser (14) that also receives the unit cell dimensions determined by the analyser (12). The structure factor analyser (14) analyses the experimental diffraction pattern using the lowest symmetry space group consistent with the crystal class determined by the cell dimension analyser (12), reducing the data to a first structure factor intensity listing and an associated covariance matrix. From this listing, the true space group (16) of the crystal structure is determined and used by the structure factor analyser (14) to generate a second structure factor intensity listing and associated covariance matrix (18). By generating this second structure factor intensity listing and associated covariance matrix (18), the total quantity of the original experimental diffraction data is significantly reduced in amount without loss of those characteristics of the data representative of the crystal structure under examination. The original data can be reduced by a factor typically equal to the number of data points in the original powder diffraction data divided by the number Bragg

reflections in the measured data range. Thus, the experimental diffraction data is not presented for analysis as a point-by-point profile, but rather in a reduced data form enabling the data to be analysed using a fitness function described in greater detail below. In performing the reduction of the original data the individual Bragg reflections (peaks) are identified and each reflection is allocated a value representative of its intensity along with one or more factors that describe the extent to which the reflection overlaps with one or more adjacent reflections.

As an example of the extent to which the powder diffraction data is reduced, typically powder diffraction data will contain between 2000 and 5000 individual data points with the reduced representation of the diffraction data the total number of data points is reduced to between 100 and 400. Hence, the original data is reduced by a factor of between 10 and 30 whilst still retaining the characteristics of the original data. Furthermore, the speed of analysis of each individual data point is increased over conventional techniques because with the original powder diffraction data the analysis of each data point includes not only the structure factor but also a factor representative of the shape of the peak. With the reduced data the analysis of each data point is reduced to only the structure factor.

Where the selected space group based on the unit cell dimensions does not prove to be the correct space group, the method described above is repeated using different space groups until the correct space group is identified.

Using the known 2D chemical formula for cimetidine (20), a co-ordinate generator (22) determines a set of internal co-ordinates (24) which completely describe the three dimensional structure of the molecule. The internal co-ordinates (24) include known data using tabulated bond lengths, bond angles and rigid torsion angles, where necessary, along with identification of unknown variables such as flexible torsion angles. When postulated values for the unknown variables are added, sufficient information is present in the internal co-ordinates to define the

conformation of an isolated theoretical cimetidine molecule.

Preferably, the only unknown factors and so the only variables to be found in the internal co-ordinates are the values of the variable torsion angles (represented by variables  $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \dots$ ). It is not essential for the bond lengths and bond angles to be held fixed and where appropriate these factors too may be varied in determining the crystal structure of the molecule. It has been found though that variation of the bond lengths and bond angles within chemically sensible bounds has a much smaller effect on the calculated diffraction data than variation of the flexible torsion angles within the structure. Thus for most purposes, acceptable results can be achieved with these factors held fixed.

Generation of the set of internal co-ordinates may alternatively be performed manually in which case the manually generated set of internal co-ordinates is input into the crystal modelling apparatus.

The output (24) of the co-ordinate generator (22) is supplied to a controller (26) that is also connected to the space group output (16) of the structure factor analyser (14). The controller (26) also includes an input (28) to enable manual setting of selected operational parameters such as the number of trial structures to be analysed in each generation, i.e. the population size. The controller (26) uses the internal co-ordinates and the space group to determine additional variables representing the location and orientation of a molecular structure in the unit cell. Preferably, the location of the molecular structure within the unit cell is defined using a single reference point in fractional co-ordinate space represented by external co-ordinates or variables ( $x, y, z$ ). The orientation of the molecule at that point may be described using Euler angles ( $\alpha, \beta, \gamma$ ). Alternatively, the orientation of the molecule may be described using a quaternion,  $q$ .

In this way the molecular crystal structure is reduced to a set of variables consisting of internal and external co-ordinates:

$\{x, y, z, \alpha, \beta, \gamma, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \dots\}$  or  $\{x, y, z, q, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \dots\}$ .

These variables are suitable for iterative mathematical processing and are more amenable to search procedures than the full complement of



individual atomic co-ordinates used in conventional techniques.

The output (30) from the controller (26) is then supplied to an iterative searching processor (32). The output (30) consists of the set of variables determined by the controller (26); the complete internal co-ordinates produced by the co-ordinate generator (22); operating parameters such as the selected size of the population to be employed in the searching procedure; any rules restricting or controlling the values which can be allocated to each of the variables; and any rules controlling the selection of survivors, the breeding and the mutation of survivors, described in greater detail below.

In the method shown in Figure 3, the iterative searching processor (32) employs a genetic algorithm to determine the correct molecular crystal structure. The above mentioned set of variables  $\{x, y, z, \alpha, \beta, \gamma, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \dots\}$  or  $\{x, y, z, q, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5 \dots\}$  are thus equated to chromosomes, with each individual variable equating to a gene. The genetic algorithm establishes certain protocols based on the concept of 'survival of the fittest', with respect to the selection of survivors, the breeding and the mutation of survivors.

Firstly, within the searching processor (32) an initial population of chromosomes is created (33) by assigning random numbers to each of the genes of each of the chromosomes. The allowable random numbers for any particular gene may be restricted in accordance with rules input from the controller (26). The selected size of this initial population depends somewhat upon the complexity of the structure under investigation, with larger population sizes typically being required for problems involving more variables. In the case of cimetidine, where seven torsion angles were allowed to vary, resulting in thirteen degrees of freedom, a population size of 150 was chosen. The fractional co-ordinates (x, y, z) and Euler angles are randomly set as real numbers normally bounded by the Euclidian normalisers of the space group. The variable torsion angles ( $\tau$ ) are typically randomly set as real numbers in the range  $0^\circ$ - $360^\circ$ .

Using the internal co-ordinates a three dimensional structure of the

trial molecule is constructed (35) for each parent in Cartesian space, and then in fractional space with respect to the crystal unit cell. Diffraction data is then determined (37) for each of the trial molecular structures and a fitness value,  $\chi^2$ , is calculated (39) for each trial structure with respect to the structure factor intensity listing and covariance matrix. The preferred fitness function employed is as follows:

$$\chi^2 = \sum_h \sum_k \{ (I_h - c|F_h|^2) (V^{-1})_{hk} (I_k - c|F_k|^2) \}$$

10 where:

$I_{h,k}$  = extracted intensity from the structure factor analyser

$V_{hk}$  = covariance matrix from the structure factor analyser

$c$  = a scale factor

$F_{h,k}$  = calculated structure factor from trial structure

15 In determining the fitness value,  $\chi^2$ , the optimal value of the scale factor  $c$  must be determined. This is preferably done by performing a conventional linear least squares analysis to determine the optimal scaling factor between the calculated structure factors and the reduced data. In addition to optimising with respect to the scale factor, the fitness value,  $\chi^2$ , may also be optimised with respect to quantities such as an overall temperature factor or a preferred orientation parameter in which case a conventional non-linear least squares analysis is preferably performed.

The fitness values for each of the chromosomes is compared to a predetermined threshold value (41) so that in the event any one of the chromosomes is less than or equal to the threshold value a solution for the molecular structure is output (43). In the event that the fitness values of all of the chromosomes exceed the threshold value, the chromosomes are then supplied (45) to a survivor selector (47). At the same time a counter (49) is increased by one so that a record of the number of generations created is maintained.

Using the fitness values obtained for each of the chromosomes, the survivor selector (47) employs a proportional selection scheme, in which

the chances of a chromosome surviving are proportional to its fitness, to select a number of survivors. Other criteria for selecting survivors may alternatively be used. For example, a tournament selection may be employed in which case two chromosomes are selected at random and  
5 compared with one another, with the fittest surviving. In particular the Boltzmann tournament may be used as it introduces an element of simulated annealing to the selection process. In addition, the selection may be elitist with the best member of the population in terms of fitness always surviving to enter the next generation.

10 Additional fitness functions may also be employed instead of, or in combination with the aforementioned fitness function, to further enhance the analysis of the trial structures. For example, simultaneous fitting of both X-ray and neutron diffraction data; use of a molecular packing function; use of an isolated molecule Lennard-Jones type calculation; use  
15 of a rotation / translation function; and use of phase information derived from direct / Patterson methods.

Although the above method is described in terms of the entire population being subject to a common selection, the population may be divided into sub-populations in which each sub-population evolves  
20 independently of the other sub-populations albeit that migration from one sub-population to another can be enabled.

The surviving chromosomes are then used to create offspring (51) by allowing the chromosomes to 'breed'. For example, individual genes from different chromosome survivors may be mixed and/or one or more of  
25 the genes in a chromosome survivor or its offspring may be mutated by random selection of a new value for the gene. Often, the population size is kept constant throughout this breeding process. The three dimensional structure of each of the offspring is then determined (35), as before, and theoretical diffraction data calculated (37).

30 The fitness ( $\chi^2$ ) is then evaluated (39) for each of the offspring and the fitness results compared (41) to the predetermined threshold value to determine whether a likely crystal structure for the molecule has been

identified. If one of the offspring chromosomes has a fitness value which is less than or equal to the threshold value, or if a predetermined maximum number of generations has been reached, then the search procedure is stopped (43). On the other hand, if the fitness functions of the  
5 chromosome offspring all exceed the threshold value and the counted number of generations is less than the maximum allowed number, then the offspring are returned for the selection of survivors (47) and for the creation of new offspring (51).

Additional rules may also be employed where appropriate to  
10 constrain the allowable values for the variables. These rules are determined by the controller (26) that may utilise data on crystal fragments stored in a memory (53). For example, the controller (26) may search through stored crystallographic databases of known crystal structure fragments related to the molecule to provide prior information about torsion  
15 angle values likely to be adopted by the structure. Such information can then be implemented either as hard limits on the allowable values the torsion angles may adopt, or as probability distributions for the torsion angles. Furthermore, fragments of the molecule may be located using Patterson methods or direct methods. For example, the location of a  
20 heavy atom may be used to anchor a molecule during the analysis by the searching processor (32). This effectively reduces the dimensionality of the problem by three as the fractional reference co-ordinates are then known.

Operation of the processor (32) in the search for the correct 3D  
25 molecular crystal structure is thus an iterative procedure with the average fitness for each generation gradually tending towards the global minimum in fitness function space. In Figure 5a, a trial cimetidine crystal structure, corresponding to a chromosome in the first generation initialised at random by the processor (32), is shown overlying the true crystal structure first  
30 shown in Figure 4. Figure 5b then shows one of the early offspring determined by the processor having a fitness value of  $\chi^2=980$ , again overlying the true crystal structure of cimetidine. In Figure 5c, a later

offspring having a fitness value of  $\chi^2=430$  is shown and the improvement in correspondence between the trial crystal structure and the actual crystal structure is immediately evident. At this point, the crystal structure could be refined using a conventional constrained Rietveld refinement. Hence,  
5 the processor (32) may be arranged so that the threshold value for the fitness function is set at around 450. This would result in the search procedure being stopped once the trial structure shown in Figure 5c had been generated, thereby enabling alternative methods to be used to refine the fine details of the trial structure. The advantage of stopping the search  
10 procedure at this point is that, usually, conventional methods will be able to refine the fine details of the structure more efficiently than the presently described method and apparatus.

Continuing with the present method, in Figure 5d an offspring having a fitness value of  $\chi^2=110$  is shown at which point the detail of the trial  
15 structure is easily refinable. Figure 6 is a graph of trial results for cimetidine using the method described above showing the fitness value of offspring with respect to the number of generations for both average fitness and the best fitness. As can be seen, a refinable structure is obtained within a few hundred generations, and an easily refinable structure is  
20 obtained around 3000 generations. This latter structure corresponds to an elapsed time of approximate 40 minutes, with the processor running on a single 175MHz R10000 Silicon Graphics™ workstation.

As further examples for the speed of this method, easily refinable structures for pyrene were determined in around 33 seconds, around 15  
25 seconds for chlorothiazide and 36 minutes for Ibuprofen, with all calculations being performed on a single 200MHz Pentium Pro™ personal computer.

The above method and apparatus may also be used with molecular structures consisting of more than one fragment. As shown in Figures 7a,  
30 7b and 7c an easily refinable structure solution for dopamine deuterobromide using neutron powder diffraction data was achieved in around only 4000 generations. This structure involves not only a dopamine

cation, but also a separate bromide anion. Using the present method and apparatus the location, orientation and conformation of the cation, and the location of the anion can be determined simultaneously.

Whilst in the examples given above the individual genes are real numbers, they could equally be represented by binary strings or integer approximations with appropriate scaling factors. Also, in the example given above the experimental diffraction data is reduced to a structure factor listing and associated covariance matrix, it will be apparent that alternative ways of reducing the total quantity of diffraction data may be employed which, although providing less faithful representations of the diffraction data, nevertheless preserve sufficient characteristics of the original diffraction data to enable a successful structure determination to be performed. For example the data relating to the correlation of reflections may be omitted from the reduced representation.

In the above example a genetic algorithm searching processor is employed to perform an iterative selection of candidate molecular crystal structures. Alternative iterative analysis processes such as simulated annealing, evolutionary strategies and neural network analysis may be used instead of the genetic algorithm. For example, using a simulated annealing process, only a single member is normally utilised and the same variables that were treated as genes by the genetic algorithm are individually adjusted by a small perturbation of their current values. If the function value ( $\chi^2$  as defined previously) is better than before, then the new values of the variables are retained. If the function value is worse, then the new values of the variables are not automatically rejected. Instead the new values may be retained if allowed by the temperature dependent Boltzmann selection protocol. In this way, 'uphill' (in terms of  $\chi^2$ ) adjustment of the variables is permissible, helping the algorithm to escape from local minima. The initial choice of the temperature is usually high to allow large 'uphill' moves if necessary, but the temperature is usually lowered in some predetermined fashion during the iterative process. One such way is a temperature reduction that cools more slowly if the  $\chi^2$

fluctuations are large.

An example employing the simulated annealing process to the determination of the crystal structure of capsaicin is shown in Figures 8 to 10. Experimental powder diffraction data for capsaicin is shown in Figure 8 with the first page of a tabulation of the reduced representation of the data produced using the above mentioned method shown in Figure 9. In Figure 10 the crystal structure obtained from powder diffraction data alone is overlaid upon the crystal structure obtained by the conventional single crystal diffraction route is shown. The experimental powder diffraction data consists of 9901 data points whilst the reduced representation of this data as exemplified in Figure 9 contains only 379 data points, in total.

Figure 9 shows the first page of the tabulated reduced representation consisting of twelve columns with columns A to C identifying the Miller indices of the Bragg reflections; column D identifying the Bragg intensity; column E an esd for the Bragg intensity; column F provides a numerical count of the separable reflections whereby identical numbers in consecutive rows indicates where reflections are so close together they cannot be separated; and columns G to L describe the percentage correlation between related reflections. In this example the powder diffraction data has been Pawley fitted using a predetermined unit cell and space group. Reflection intensities have been allowed to vary as variables in a least squares fit to the diffraction data. The result is a file containing estimates of the individual reflection intensities. However, where reflections lie very close together in 20 they may be treated as a clump. Where the reflections are close, but not close enough to be treated as a clump, the correlation terms provide information on how the respective intensities are related. Thus, the first reflection at line 1:

1 0 0 10.047 0.1106 1 0 0 0 0 0 0

is an isolated reflection whereas at lines 18 and 19 two correlated reflections are identified. At lines 41 and 42 two reflections are identified that lie so close together that they are treated as a single variable intensity. With the reduced representation the diffraction data has been effectively

condensed into the data contained in columns D and E in combination with non-zero elements of columns G to L.

The close agreement between the two structures shown in Figure 10 demonstrates one characteristic of the method described above - the ability of the simulated annealing method to 'fine tune' the structure, generally finding a solution very close to the global minimum in  $\chi^2$  space. This structure solution, which involved the optimisation of 16 degrees of freedom (10 internal, 6 external) took approximately 40 minutes to execute on a DEC Alphastation 500/500.

In another analysis, the molecule under study is said to possess a total energy equal to the sum of its potential energy (proportional to the current  $\chi^2$ ) and some kinetic energy applied to start the molecule moving over the  $\chi^2$  hypersurface. The gradient of the  $\chi^2$  hypersurface with respect to each of the degrees of freedom is calculated and it is these gradients that control the resultant trajectory of the molecule and thus the way in which it explores the hypersurface. Lowering the temperature of the system in a way analogous to that already described gradually decreases the kinetic energy of the system and that in turn leads to the molecular configuration conforming to the global minimum of the  $\chi^2$  hypersurface.

In a further adaption of the method, regardless of the global optimisation strategy used, both local and semi-global optimisation methods (e.g. Newton-Raphson, simplex) can be invoked when the  $\chi^2$  value reaches some predetermined value that is anticipated to be in the vicinity of the global minimum, thus providing accelerated convergence.

With the method and apparatus described above, molecular crystal structures may be solved from powder diffraction data alone. Definition of the molecular fragments in terms of internal co-ordinates means that for a single molecular fragment, problem complexity scales with the number of variable torsion angles rather than with the number of atoms in the fragment. Thus, complex structures can be represented by quite short chromosomes and solved relatively easily. The simple description of



molecular geometry employed, together with the genetic algorithm / simulated annealing analyses and the specified fitness function has thus been shown to be particularly powerful in determining crystal structures from powder diffraction data in a relatively short time frame.

- 5           To assist in an understanding of the invention, the method has been described with reference to functional, i.e. analyser/processor units. It will of course be apparent that in practice the method is implemented as a program on a computer. Indeed, one of the advantages of this method is that the program can be implemented on a number of different computer
- 10 architectures, including personal computers and a network of personal computers/workstations acting as a parallel computer.

## CLAIMS

1. A method for determining molecular crystal structures from powder  
diffraction data comprising the steps of: generating a reduced  
5 representation of the powder diffraction pattern in dependence on a  
predetermined unit cell and space group of the molecule under  
examination in which the total quantity of diffraction data is  
significantly reduced whilst maintaining the characteristics of the  
diffraction data that are representative of the crystal structure under  
10 examination; determining a set of variables for describing trial  
molecular structures, derived from predetermined internal co-  
ordinates and said space group; assigning values to said variables  
thereby creating a population of trial structures each defined by a  
unique set of values for said variables; calculating a fitness for each  
15 trial structure with respect to the reduced representation of the  
powder diffraction pattern; determining whether any one of the  
calculated fitnesses is less than or equal to a predetermined  
threshold; where none of the calculated fitnesses is less than or  
equal to the threshold value, selecting at least one survivor from the  
20 population of trial structures, altering the values of the variables of at  
least one of the survivors in accordance with one or more  
predetermined rules, calculating the fitnesses of the new trial  
structures; and repeating the steps of selecting survivors, altering  
the values of the variables and calculating the fitnesses of the new  
25 trial structures until at least one of the calculated fitnesses is less  
than or equal to the threshold value, and where at least one of the  
calculated fitnesses is less than or equal to the threshold, outputting  
at least one trial molecular crystal structure represented by the  
successful sets of values.  
30
2. A method as claimed in claim 1, wherein the reduced representation  
includes single values representative of the intensity of each

reflection in the powder diffraction data and one or more factors representative of the extent to which adjacent reflections overlap.

3. The method as claimed in claim 2, wherein the reduced  
5 representation consists of a structure factor intensity listing and associated covariance matrix.
4. The method as claimed in either of claims 2 or 3, wherein the total  
10 data in the reduced representation is reduced by a factor substantially equal to the number of data points in the original powder diffraction data divided by the number Bragg reflections in the measured data range.
5. The method as claimed in either of claims 3 or 4, wherein the fitness  
15  $\chi^2$  of each of the trial structures is determined using the following function:

$$\chi^2 = \sum_h \sum_k \{ (I_h - c|F_h|^2) (V^{-1})_{hk} (I_k - c|F_k|^2) \}$$

20 where:

$I_{h,k}$  = extracted intensity

$V_{hk}$  = covariance matrix

$c$  = a scale factor

$F_{h,k}$  = calculated structure factor from trial structure

25

6. The method as claimed in any one of the preceding claims, wherein  
the set of variables consists of three co-ordinates representative of  
the location of the molecule within the unit cell and three  
independent co-ordinates representative of the orientation of the  
30 molecule within the unit cell.

7. The method as claimed in claim 6, wherein the set of variables

includes one or more co-ordinates representative of variable torsion angles, bond angles or bond lengths.

- 5 8. The method as claimed in any one of the preceding claims including the step of determining the unit cell and space group for the molecule under examination.
9. The method as claimed in any one of the preceding claims including the step of determining the set of internal co-ordinates.
- 10 10. The method as claimed in any one of the preceding claims, further including the step of monitoring the number of iterations in which new trial structures are generated and halting the method and outputting the trial crystal structure with the best calculated fitness after completion of a predetermined number of iterations.
- 15 11. The method as claimed in any one of the preceding claims, wherein the selection of survivors and the alteration of the values of the variables is based on a simulated annealing procedure.
- 20 12. Apparatus for determining molecular crystal structures comprising a structure factor analyser for generating from experimental powder diffraction data for the molecule under examination a reduced representation of the powder diffraction pattern based on a predetermined unit cell and space group in which the total quantity of diffraction data is significantly reduced whilst the characteristics of the diffraction data representative of the crystal structure under examination are maintained; a controller for determining a set of variables for describing trial molecular structures, derived from  
25  
30 predetermined internal co-ordinates and said space group; a searching processor for creating a population of trial structures each defined by a unique set of values for said variables said searching

processor including a fitness analyser for calculating a fitness for each trial structure with respect to the reduced representation of the powder diffraction pattern, a thresholding device for determining whether any one of the calculated fitnesses is less than or equal to a predetermined threshold, a survivor selector for selecting at least one survivor from the population of trial structures, a variable adjustment device for altering the values of the variables of at least one of the survivors and output means for outputting the one or more trial molecular crystal structures having calculated fitnesses less than or equal to the threshold value.

13. Apparatus as claimed in claim 12, wherein the reduced representation includes single values representative of the intensity of each reflection in the powder diffraction data and one or more factors representative of the extent to which adjacent reflections overlap.
14. Apparatus as claimed in claim 13, wherein the structure factor analyser generates a reduced representation consists of a structure factor intensity listing and associated covariance matrix.
15. The method as claimed in either of claims 13 or 14, wherein the total data in the reduced representation is reduced by a factor substantially equal to the number of data points in the original powder diffraction data divided by the number Bragg reflections in the measured data range.
16. Apparatus as claimed in either of claims 14 or 15, wherein the fitness analyser determines the fitness  $\chi^2$  of each of the trial structures using the following function:

$$\chi^2 = \sum_h \sum_k \{ (I_h - c|F_h|^2) (V^{-1})_{hk} (I_k - c|F_k|^2) \}$$

where:

$I_{h,k}$  = extracted intensity from the structure factor analyser

$V_{hk}$  = covariance matrix from the structure factor analyser

$c$  = a scale factor

5  $F_{h,k}$  = calculated structure factor from trial structure

17. Apparatus as claimed in any one of claims 12 to 16, wherein  
controller determines a set of variables consists of three co-  
ordinates representative of the location of the molecule within the  
10 unit cell and three independent co-ordinates representative of the  
orientation of the molecule within the unit.
18. Apparatus as claimed in any one of claims 12 to 17, wherein the  
structure factor analyser additionally determines the unit cell and  
15 space group for the molecule under examination.
19. Apparatus as claimed in any one of claims 12 to 18 including a co-  
ordinate generator for determining the set of internal co-ordinates.
- 20 20. Apparatus as claimed in any one of claims 12 to 19, further  
including a counter for monitoring the number of iterations of new  
trial structures.

Fig 1.

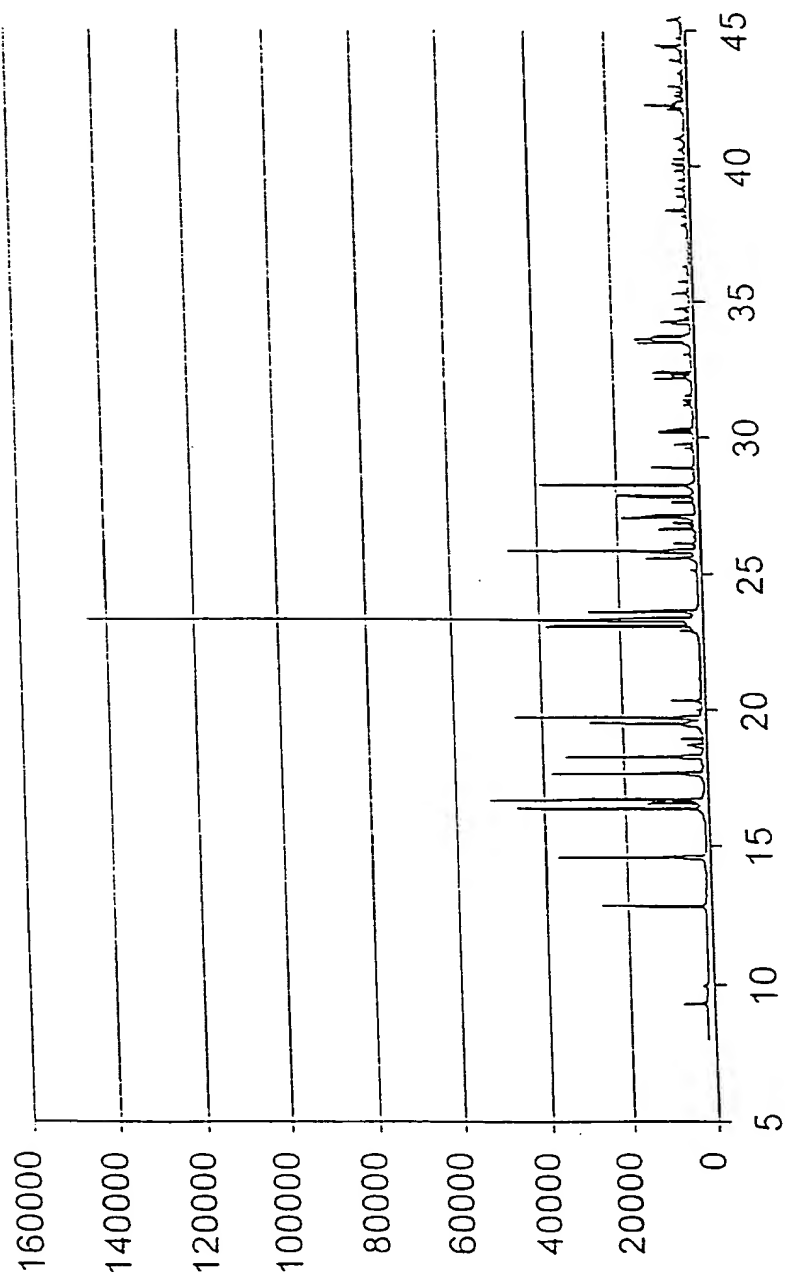
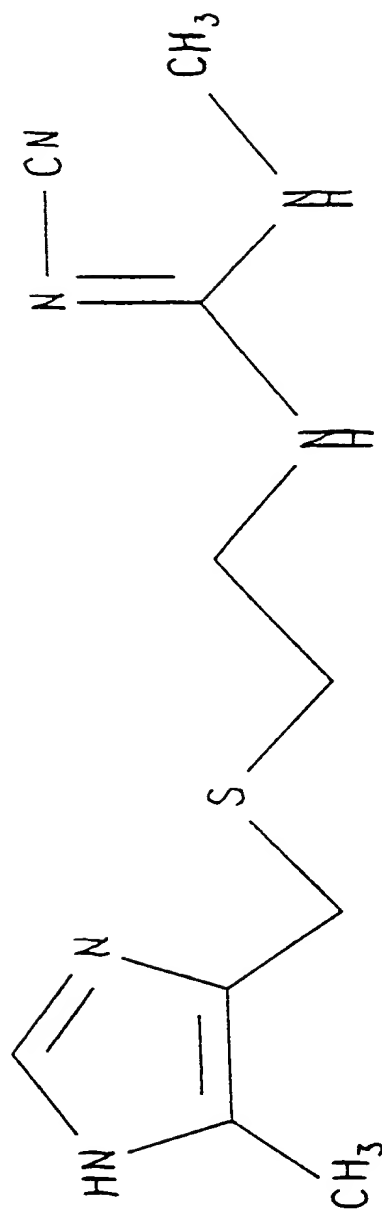


Fig 2.





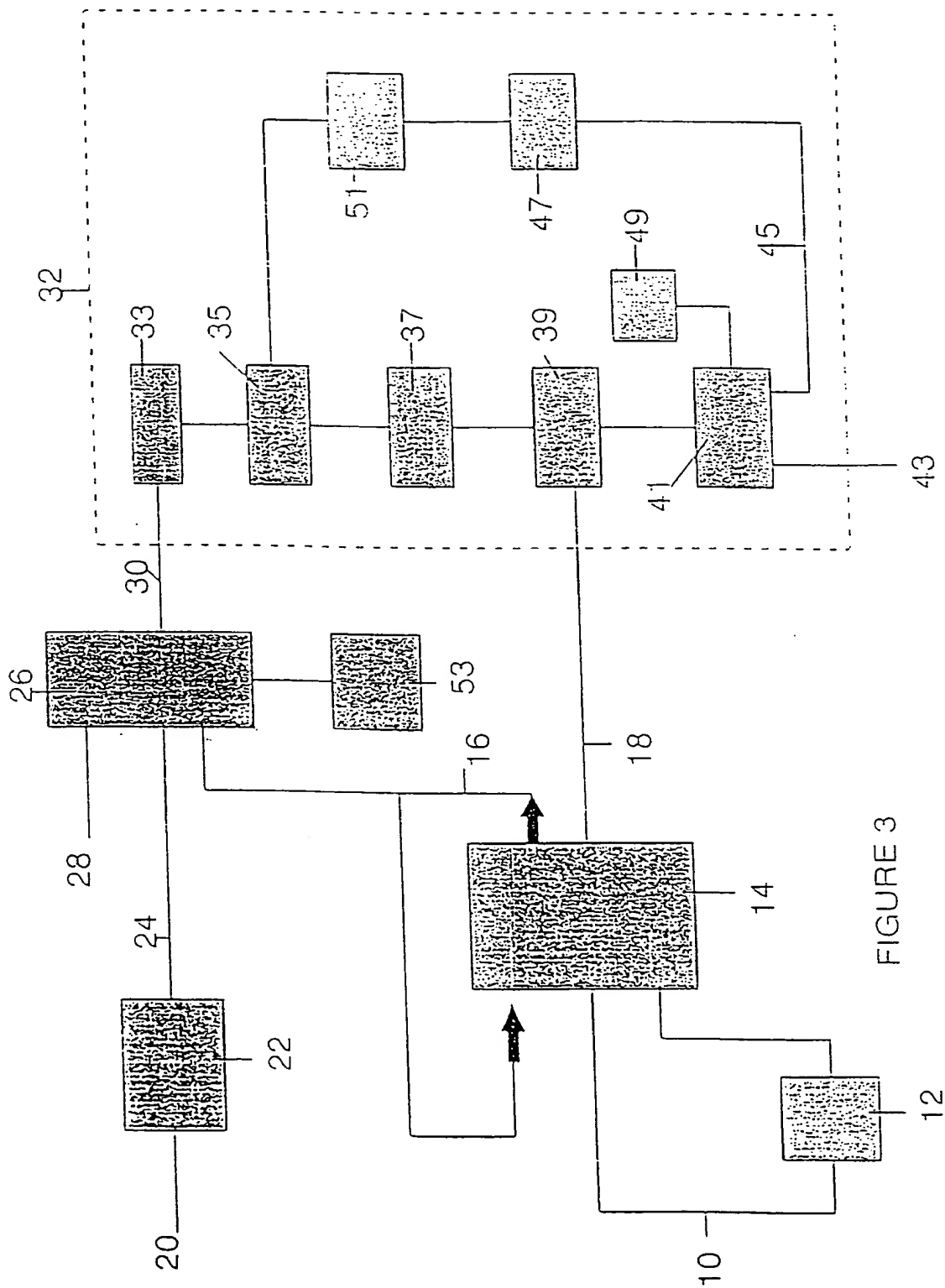


Fig 4.

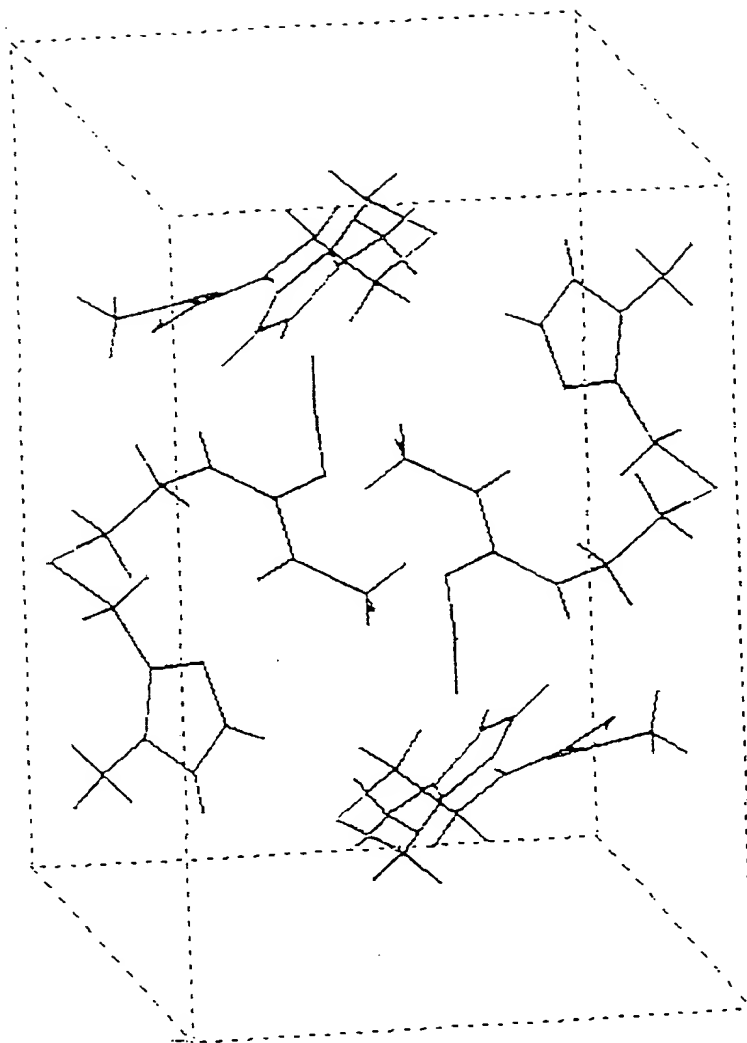
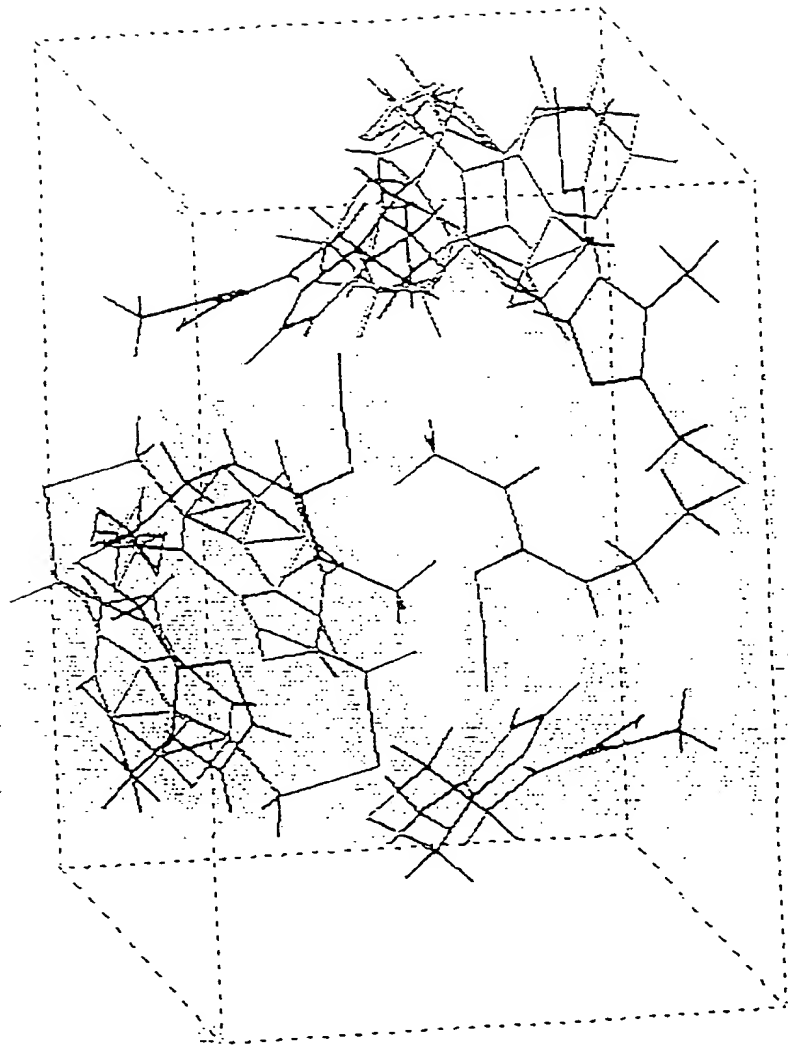
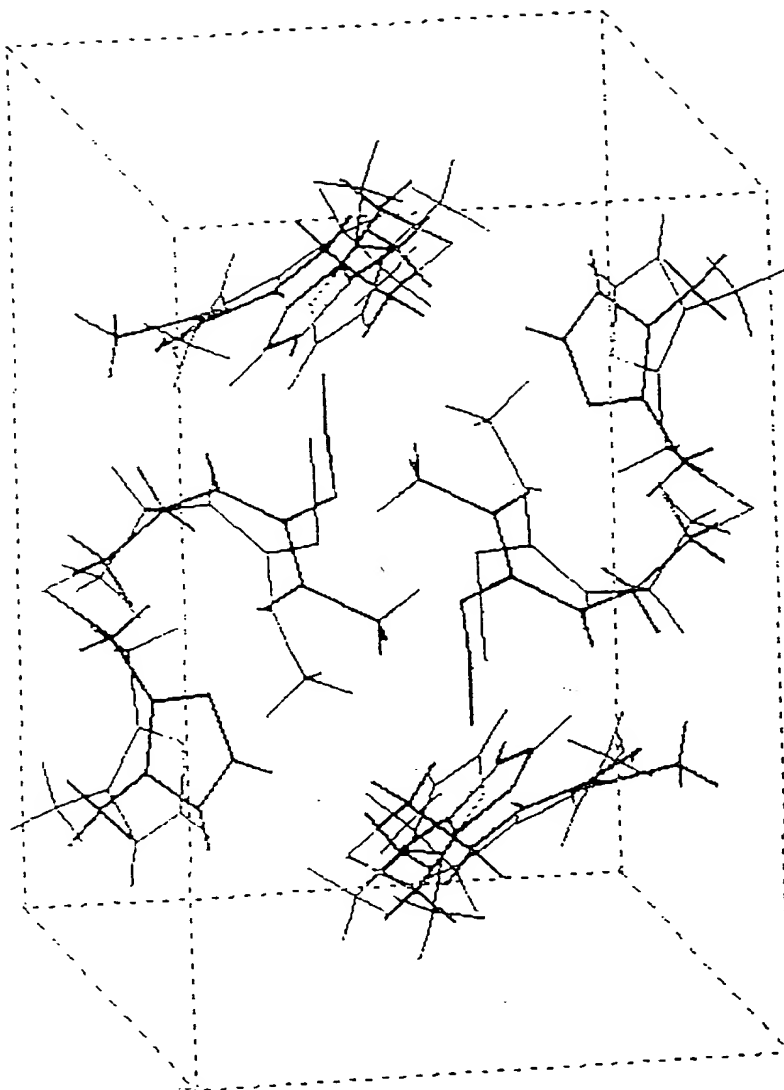


Fig 5a.



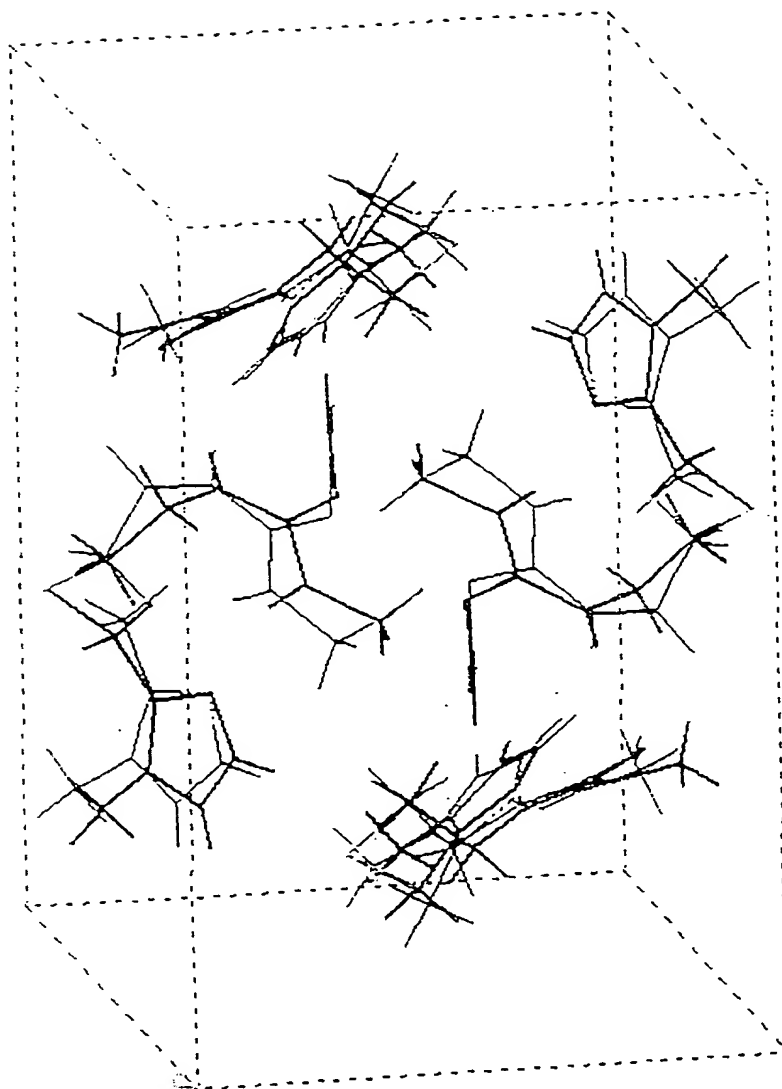
6/15

Fig 5b.



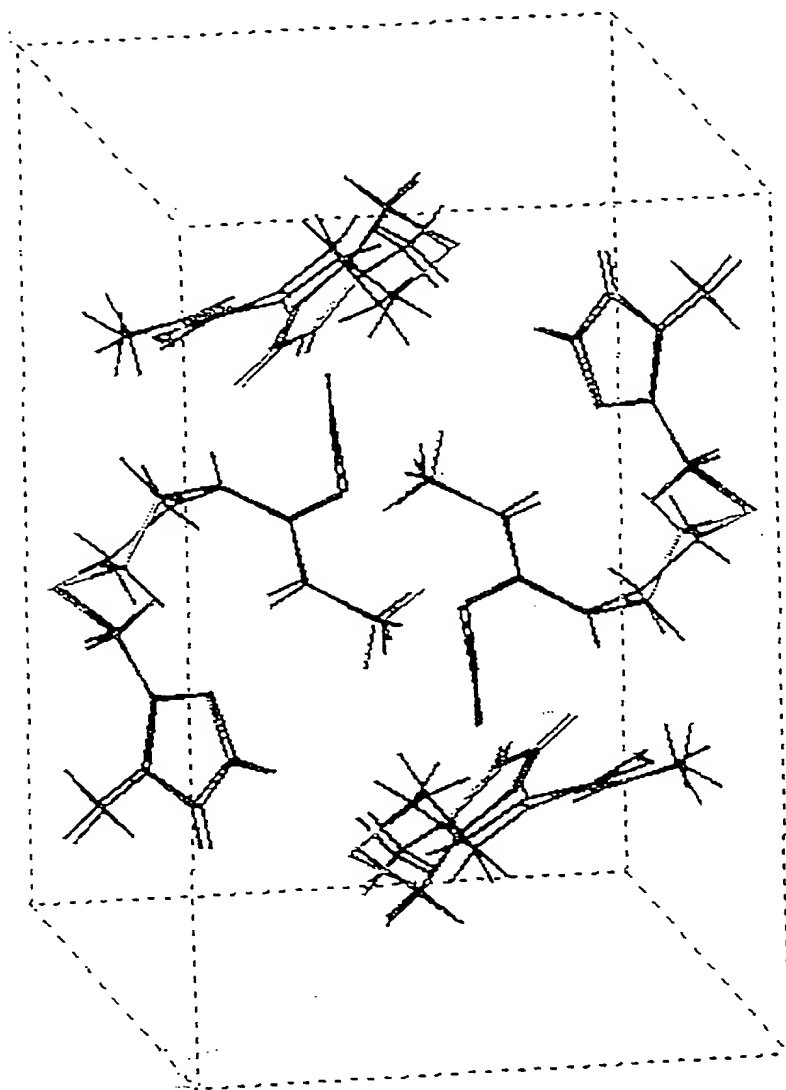
7/15

Fig 5c.



8/15

Fig 5d.



9/15

Fig 6.

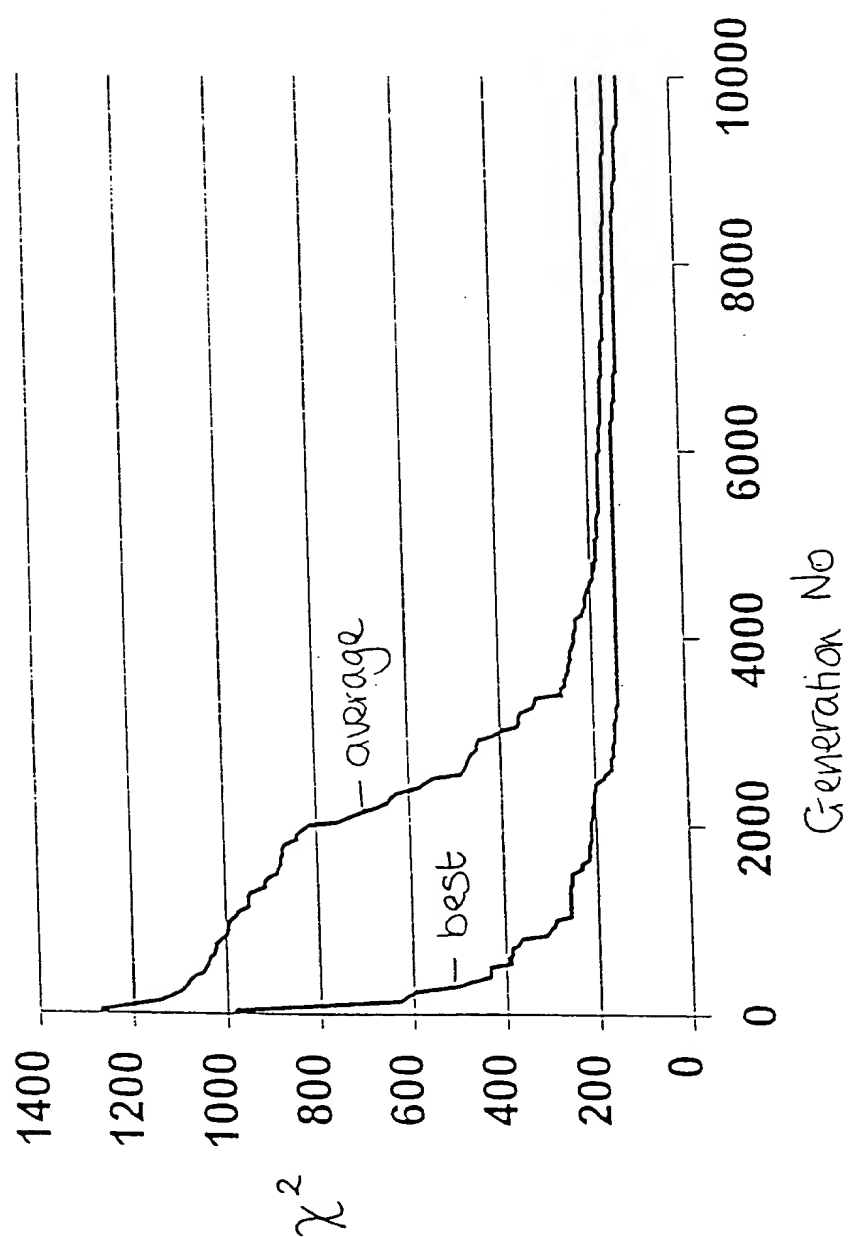
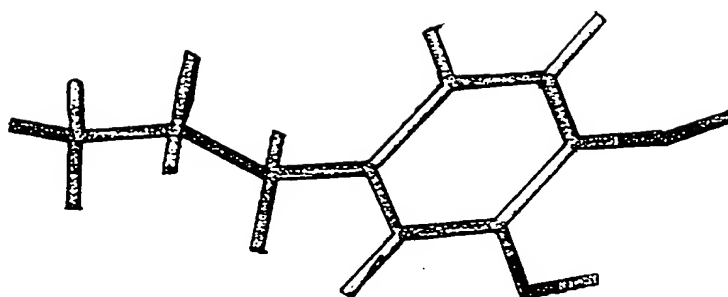


Fig 7a.





11/15

Fig 7b.

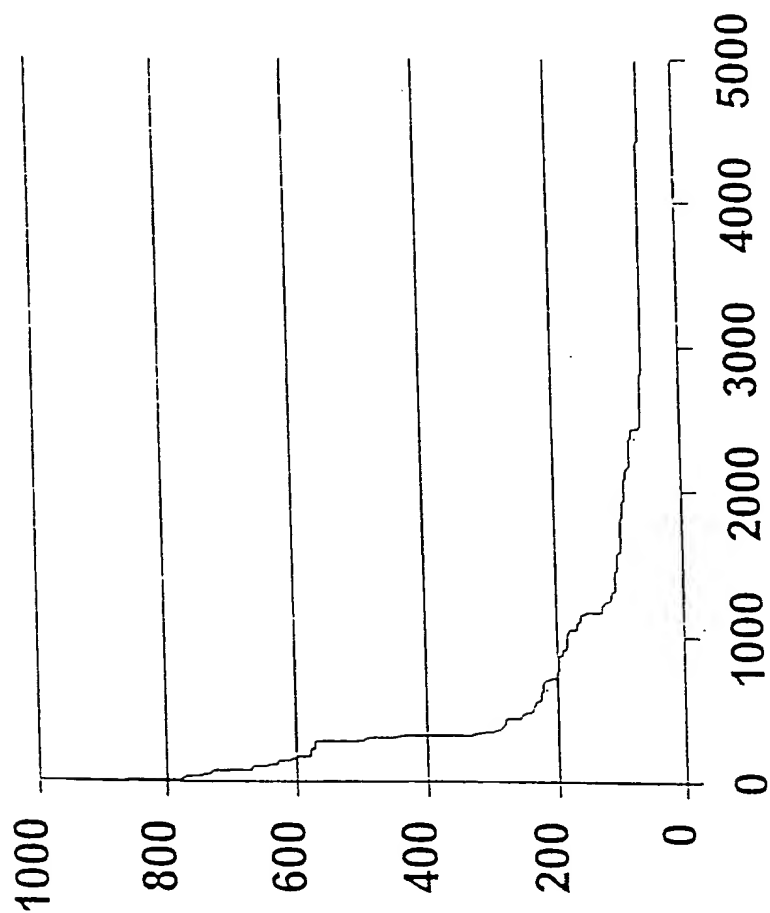
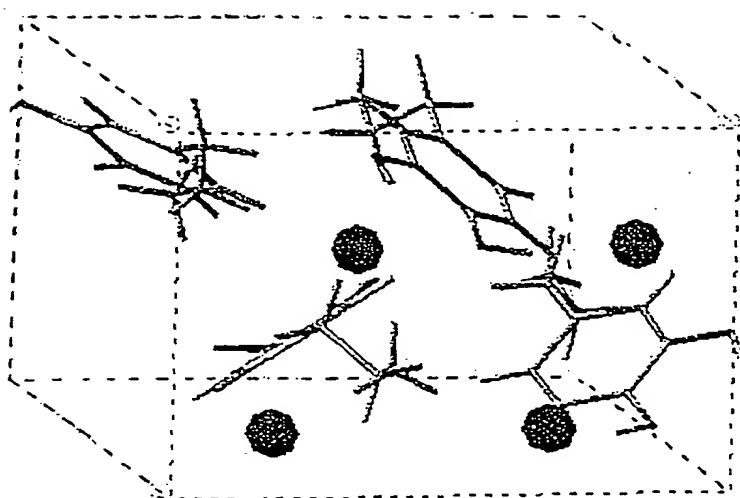
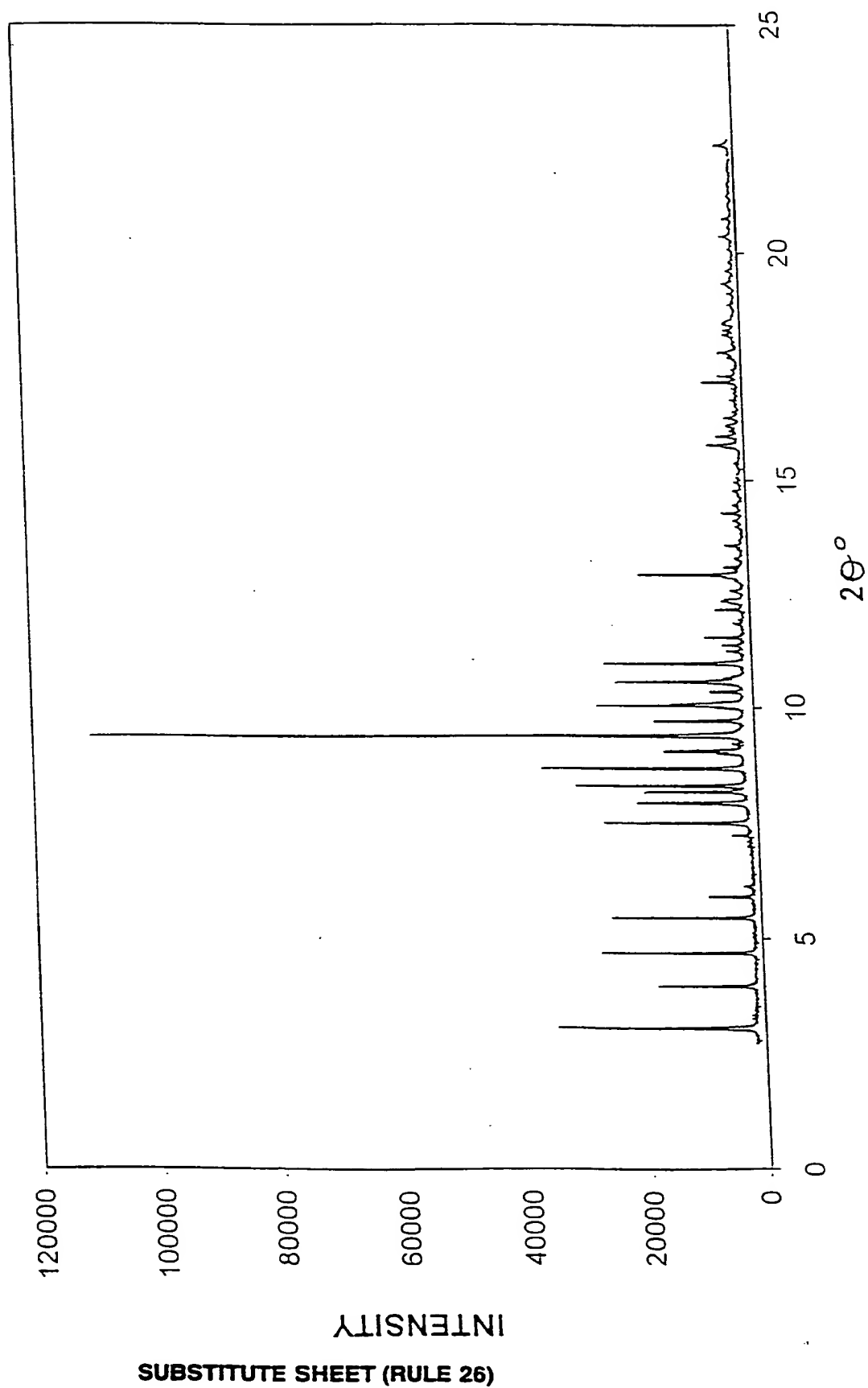


Fig 7c.



13/15

FIGURE 8



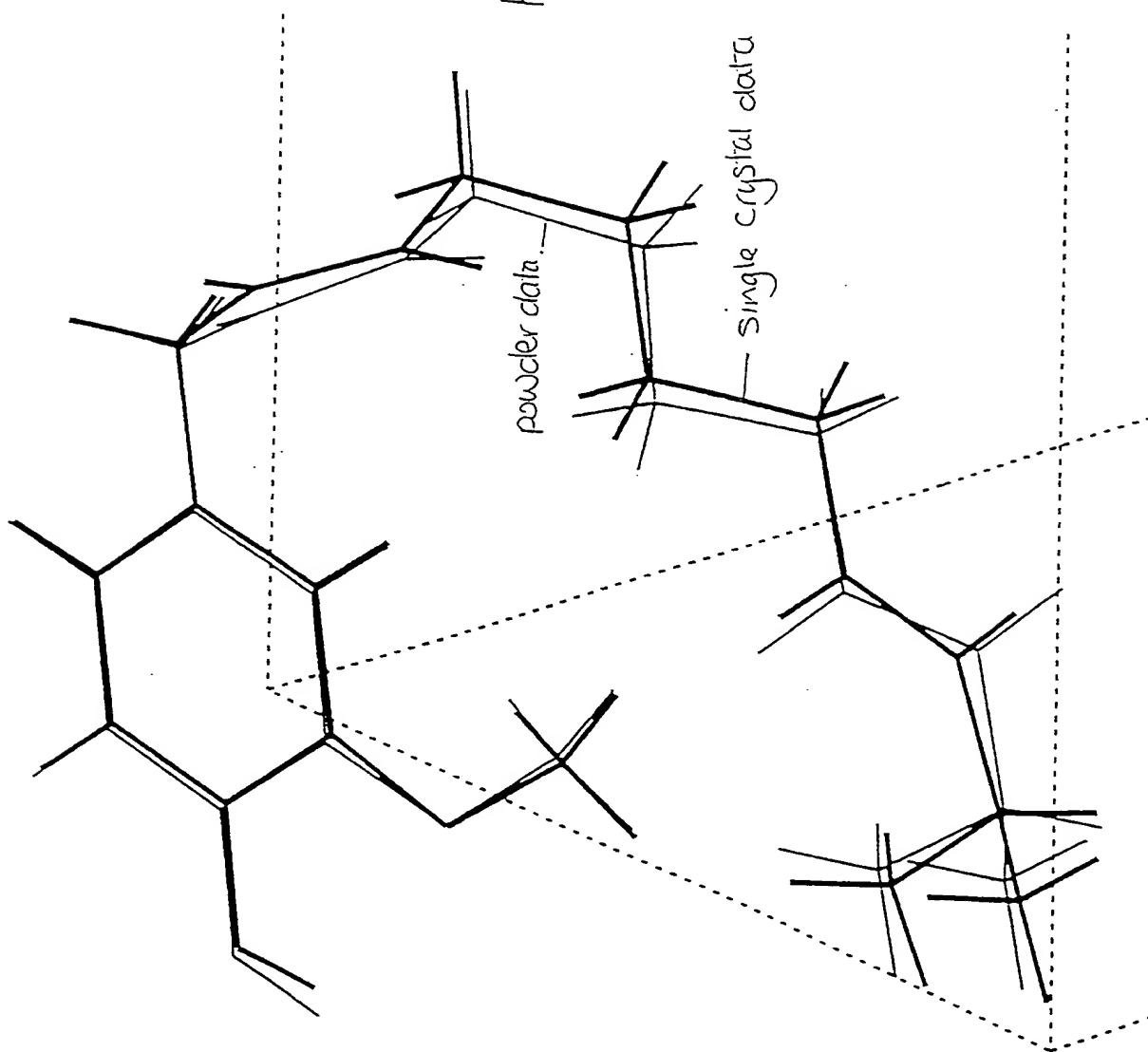
14/15

FIGURE 9

	A	B	C	D	E	F	G	H	I	J	K	L
1	1	0	0	10.047	0.1106	1	0	0	0	0	0	0
2	-1	1	0	3.637	0.0617	2	0	0	0	0	0	0
3	0	1	1	5.754	0.0874	3	0	0	0	0	0	0
4	0	2	0	0.321	0.0666	4	0	0	0	0	0	0
5	-1	1	1	8.983	0.1116	5	0	0	0	0	0	0
6	1	1	1	0.294	0.0432	6	0	0	0	0	0	0
7	-1	2	0	3.956	0.0916	7	0	0	0	0	0	0
8	2	0	0	2.070	0.1117	8	0	0	0	0	0	0
9	0	2	1	-0.025	0.0446	9	0	0	0	0	0	0
10	-2	1	0	0.005	0.0493	10	0	0	0	0	0	0
11	-1	2	1	0.375	0.0660	11	0	0	0	0	0	0
12	1	2	1	2.156	0.0959	12	0	0	0	0	0	0
13	-2	1	1	21.121	0.2266	13	0	0	0	0	0	0
14	0	0	2	10.592	0.4909	14	-55	34	0	0	0	0
15	2	1	1	9.634	0.6372	15	-88	0	0	0	0	0
16	-2	2	0	9.055	0.5427	16	0	0	0	0	0	0
17	-1	3	0	17.781	0.2205	17	0	0	0	0	0	0
18	-1	0	2	8.314	0.5049	18	-57	0	0	0	0	0
19	0	1	2	25.368	0.3353	19	0	0	0	0	0	0
20	0	3	1	0.385	0.0972	20	0	0	0	0	0	0
21	-1	1	2	2.992	0.2855	21	-59	0	0	0	0	0
22	1	0	2	23.776	0.2748	22	0	0	0	0	0	0
23	-2	2	1	23.776	0.2748	23	0	0	0	0	0	0
24	-1	3	1	5.447	0.1809	24	0	0	0	0	0	0
25	1	1	2	10.351	0.3342	25	-59	0	0	0	0	0
26	2	2	1	16.338	0.3506	26	0	0	0	0	0	0
27	1	3	1	0.341	0.1841	27	-69	0	0	0	0	0
28	3	0	0	4.652	0.3947	28	0	0	0	0	0	0
29	0	2	2	137.393	0.6686	29	0	0	0	0	0	0
30	-3	1	0	1.685	0.1391	30	0	0	0	0	0	0
31	-2	0	2	0.293	0.3221	31	-28	0	0	0	0	0
32	-1	2	2	21.709	0.3157	32	-23	0	0	0	0	0
33	-2	3	0	0.647	0.1616	33	0	0	0	0	0	0
34	-2	1	2	6.018	0.2122	34	0	0	0	0	0	0
35	1	2	2	34.450	0.4173	35	-39	0	0	0	0	0
36	-3	1	1	14.637	0.3554	36	-31	0	0	0	0	0
37	0	4	0	12.411	0.4668	37	0	0	0	0	0	0
38	2	0	2	17.335	0.5226	38	-53	0	0	0	0	0
39	-2	3	1	1.605	0.2200	39	0	0	0	0	0	0
40	-3	2	0	2.854	0.1700	40	0	0	0	0	0	0
41	-1	4	0	22.869	0.2070	41	0	0	0	0	0	0
42	3	1	1	22.869	0.2070	42	0	0	0	0	0	0
43	2	1	2	4.788	0.2129	43	-27	0	0	0	0	0
44	2	3	1	1.945	0.1824	44	0	0	0	0	0	0
45	0	4	1	-0.298	0.1358	45	0	0	0	0	0	0
46	-2	2	2	0.088	0.1723	46	0	0	0	0	0	0
47	0	3	2	44.292	0.5083	47	-48	0	0	0	0	0
48	-3	2	1	3.714	0.3270	48	0	0	0	0	0	0
49	-1	4	1	1.319	0.1982	49	-44	0	0	0	0	0
50	-1	3	2	4.787	0.2325	50	0	0	0	0	0	0
51	1	4	1	7.086	0.2216	51	0	0	0	0	0	0
52	3	2	1	-0.509	0.1419	52	0	0	0	0	0	0

15/15

Figure 10



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/02316

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G01N23/20

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	K. SHANKLAND: "Crystal structure determination from powder diffraction data by the application of a genetic algorithm." ZEITSCHRIFT FÜR PHYSIK, vol. 212, no. 8, 8 August 1997, pages 550-552, XP002087730 see page 551 - page 1 ---	1, 12
P, X	B. M. KARIUKI ET AL.: "The application of a genetic algorithm for solving crystal structures from powder diffraction data." CHEMICAL PHYSICS LETTERS, vol. 280, no. 3, 4, 5 December 1997, pages 189-195, XP002087731 see page 190, last paragraph - page 191; figure 1 -----	1, 12

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

14 December 1998

Date of mailing of the international search report

29/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hulne, S

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>G01N 23/20</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 99/06824</b> <b>(43) International Publication Date:</b> 11 February 1999 (11.02.99)
<b>(21) International Application Number:</b> PCT/GB98/02316 <b>(22) International Filing Date:</b> 31 July 1998 (31.07.98)  <b>(30) Priority Data:</b> 9716278.8                      31 July 1997 (31.07.97)                      GB  <b>(71) Applicant (for all designated States except US):</b> COUNCIL FOR THE CENTRAL LABORATORY OF THE RESEARCH COUNCILS [GB/GB]; Chilton, Didcot, Oxfordshire OX11 0QX (GB).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DAVID, William, Ian, Fraser [GB/GB]; 24 Appleford Drive, Abingdon, Oxfordshire OX14 2BZ (GB). SHANKLAND, Kenneth [GB/GB]; 83a The Vineyard, Abingdon, Oxford OX14 3PG (GB).  <b>(74) Agent:</b> PERKINS, Sarah; Stevens, Hewlett & Perkins, 1 Serjeants' Inn, Fleet Street, London EC4Y 1LL (GB).		<b>(81) Designated States:</b> JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> METHOD AND APPARATUS FOR DETERMINING MOLECULAR CRYSTAL STRUCTURES  <b>(57) Abstract</b>  An improved method and apparatus for determining molecular crystal structures is provided which enables molecular crystal structures to be identified using only powder diffraction data, that is considerably faster than conventional crystal modelling techniques. With the improved crystal modelling method and apparatus trial molecular crystal structures are reduced to a unique set of variables based on co-ordinates representative of the location and orientation of the molecule and where appropriate at least one co-ordinate representative of a torsion angle, bond length or bond angle. In addition, the total quantity of experimental powder diffraction data is reduced to provide a reduced representation of the diffraction data in the form of a structure factor intensity listing and covariance matrix. Trial crystal structures are postulated and each is defined using the set of variables which are used in determining a fitness of each trial structure with respect to the reduced representation of the experimental data. A trial crystal structure is output as an accurate representation of the actual crystal structure, when the fitness value for the trial structure is less than or equal to a predetermined threshold value. With the improved crystal modelling method and apparatus, identification of complex molecular crystal structures can be performed in seconds or minutes using the current generation of conventional personal computers or workstations as opposed to the hours and often days required with conventional techniques.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	L Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/02316

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 6 G01N23/20

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	K. SHANKLAND: "Crystal structure determination from powder diffraction data by the application of a genetic algorithm." ZEITSCHRIFT FÜR PHYSIK, vol. 212, no. 8, 8 August 1997, pages 550-552, XP002087730 see page 551 - page 1	1,12
P,X	B. M. KARIUKI ET AL.: "The application of a genetic algorithm for solving crystal structures from powder diffraction data." CHEMICAL PHYSICS LETTERS, vol. 280, no. 3,4, 5 December 1997, pages 189-195, XP002087731 see page 190, last paragraph - page 191; figure 1	1,12

☐ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

\* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

14 December 1998

Date of mailing of the international search report

29/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hulne, S

# PATENT COOPERATION TREATY

# PCT

## INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference <b>SP/1796 PCT</b>	<b>FOR FURTHER ACTION</b> <small>see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.</small>	
International application No. <b>PCT/GB 98/ 02316</b>	International filing date ( <i>day/month/year</i> ) <b>31/07/1998</b>	(Earliest) Priority Date ( <i>day/month/year</i> ) <b>31/07/1997</b>
Applicant  <b>COUNCIL FOR THE CENTRAL LABORATORY OF THE ...et al.</b>		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 2 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. ☐ Certain claims were found unsearchable (see Box I).
2. ☐ Unity of invention is lacking (see Box II).
3. ☐ The international application contains disclosure of a **nucleotide and/or amino acid sequence listing** and the international search was carried out on the basis of the sequence listing
 

☐ filed with the international application.  
☐ furnished by the applicant separately from the international application,  
     ☐ but not accompanied by a statement to the effect that it did not include matter going beyond the disclosure in the international application as filed.  
  
☐ Transcribed by this Authority
4. With regard to the title, ☒ the text is approved as submitted by the applicant  
     ☐ the text has been established by this Authority to read as follows:
5. With regard to the abstract, ☒ the text is approved as submitted by the applicant  
     ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this International Search Report, submit comments to this Authority.
6. The figure of the **drawings** to be published with the abstract is:
 

Figure No. \_\_\_\_\_ ☐ as suggested by the applicant. ☒ None of the figures.  
☐ because the applicant failed to suggest a figure.  
☐ because this figure better characterizes the invention.

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 98/02316

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G01N23/20

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	K. SHANKLAND: "Crystal structure determination from powder diffraction data by the application of a genetic algorithm." ZEITSCHRIFT FÜR PHYSIK, vol. 212, no. 8, 8 August 1997, pages 550-552, XP002087730 see page 551 - page 1	1,12
P,X	B. M. KARIUKI ET AL.: "The application of a genetic algorithm for solving crystal structures from powder diffraction data." CHEMICAL PHYSICS LETTERS, vol. 280, no. 3,4, 5 December 1997, pages 189-195, XP002087731 see page 190, last paragraph - page 191; figure 1	1,12



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&amp;" document member of the same patent family

Date of the actual completion of the international search

14 December 1998

Date of mailing of the international search report

29/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hulne, S